HW4\_Patrick Neyland

Patrick Neyland

2022-09-13

## Question 1

### Part i

Interpret the coefficient on log(dist). Is the sign of this estimate what you expect it to be?

This model is an example of one that has undergone a log-log transformation—A 1.000 percent increase in distance from the garbage incinerator, holding all else constant, will result in 0.132 percent in price on average. This coefficient is expected—on average, people want to live far away from a presumably stinky and ugly garbage handling facility.

### Part ii

What other factors about a house affect its price? Might these be correlated with distance from the incinerator?

Proximity to grocery stores, schools, and rec centers will also affect the price of a home—these specific factors might be correlated with distance from the incinerator. I would imagine that most businesses would also want to be located far away from any garbage handling facilities.

### Part iii

Do you think simple regression provides an unbiased estimator of the ceteris paribus elasticity of price with respect to dist? (Think about the city’s decision on where to put the incinerator. Hint: which of the SLR assumptions is/are violated?)

## Question 2

x<- 1:30  
y <- c(1)  
for (i in x){  
 y[i] <- 1.1\*tail(y,1)  
}  
y

[1] 1.100000 1.210000 1.331000 1.464100 1.610510 1.771561 1.948717  
 [8] 2.143589 2.357948 2.593742 2.853117 3.138428 3.452271 3.797498  
[15] 4.177248 4.594973 5.054470 5.559917 6.115909 6.727500 7.400250  
[22] 8.140275 8.954302 9.849733 10.834706 11.918177 13.109994 14.420994  
[29] 15.863093 17.449402

library(jtools)  
#summ(lm(y~x))  
  
#stargazer(lm(y~x), lm(log(y)~x), type="text")

## Question 3

### Part i

Find the average participation rate and the average match rate in the sample of plans.

mean(k401k$prate)

[1] 87.36291

mean(k401k$mrate)

[1] 0.7315124

The average participation rate is 87.36291 percent. The average match rate is 73.15124 percent.

### Part ii

Now, estimate the simple regression equation [prate = ^ 0 + ^ 1mrate and report the results along with the sample size and R-squared

model3\_2 <- lm(prate~mrate,k401k)  
#stargazer(reg\_3\_2, type = "text")

### Part iii

Interpret the intercept in your equation. Interpret the coefficient on . The intercept is the is the average participation rate when is zero. The coefficient of shows that on average, a one unit increase in , holding all else constant, will result in a 5.861 percent increase in participation rate.

### Part iv

prate <- function(mrate) {83.075 + 5.861\*mrate}  
prate(3.5)

[1] 103.5885

The predicted when is 103.5885 percent. This is not a reasonable prediction because in reality, cannot exceed 100 percent.

### Part v

7.5 percent of the variation in prate is explain by mrate. This does not seem like a lot to me.

## Question 4

### Part i

Do you think each additional dollar spent has the same effect on the pass rate, or does a diminishing effect seem more appropriate? Explain.

I think a diminishing effect seems more appropriate. At the beginning each dollar makes a really big difference—buying books, pencils, paper, and basic computers open tremendous opportunities for the students to learn and develop their skills. However, spending an additional $2,000 per student on a fancier computer lab will likely have little impact on test scores because it would not offer significantly more practical functionality than the normal computer lab.

model4\_2 <- lm(math10~log(expend), data = meap93)  
summary(model4\_2)

Call:  
lm(formula = math10 ~ log(expend), data = meap93)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-22.343 -7.100 -0.914 6.148 39.093   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -69.341 26.530 -2.614 0.009290 \*\*   
log(expend) 11.164 3.169 3.523 0.000475 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 10.35 on 406 degrees of freedom  
Multiple R-squared: 0.02966, Adjusted R-squared: 0.02727   
F-statistic: 12.41 on 1 and 406 DF, p-value: 0.0004752

## Question 5

### Part i

What is the average gift in the sample of 4,268 people (in Dutch guilders)? What percentage of people gave no gift?

mean(charity$gift)

[1] 7.44447

tb1 <- table(charity$respond)  
prop.table(tb1)

0 1   
0.6000469 0.3999531

The average gift in the sample is worth 7.44 Dutch guilders. The percentage of people who gave no gift is 39.99531 percent.

### Part ii

mean(charity$mailsyear)

[1] 2.049555

min(charity$mailsyear)

[1] 0.25

max(charity$mailsyear)

[1] 3.5

The average number of mailings per year is 2.05. The minimum mailings per year is 0.25. The maximum mailings per year is 3.5.

### Part iii

model5\_3 <- lm(gift~mailsyear, data = charity)  
stargazer(model5\_3, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 gift   
-----------------------------------------------  
mailsyear 2.650\*\*\*   
 (0.343)   
   
Constant 2.014\*\*\*   
 (0.739)   
   
-----------------------------------------------  
Observations 4,268   
R2 0.014   
Adjusted R2 0.014   
Residual Std. Error 14.960 (df = 4266)   
F Statistic 59.649\*\*\* (df = 1; 4266)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Part iv

Interpret the slope coefficient. If each mailing costs one guilder, is the charity expected to make a net gain on each mailing? Does this mean the charity makes a net gain on every mailing? Explain.

On average, holding all else constant, if the number of mailings per year increases by 1 unit, the amount of gift will increase by 2.65 guilders.

## Question 6

### Part i

count(catholic)

n  
1 7430

mean(catholic$math12)

[1] 52.13362

sd(catholic$math12)

[1] 9.459117

mean(catholic$read12)

[1] 51.7724

sd(catholic$read12)

[1] 9.407761

The sample size 7430 students. The average math score is 52.13362. The standard deviation of math scores is 9.459117. The average reading score is 51.7724. The standard deviation of reading scores is 9.407761.

### Part ii